deeplearning for nlp

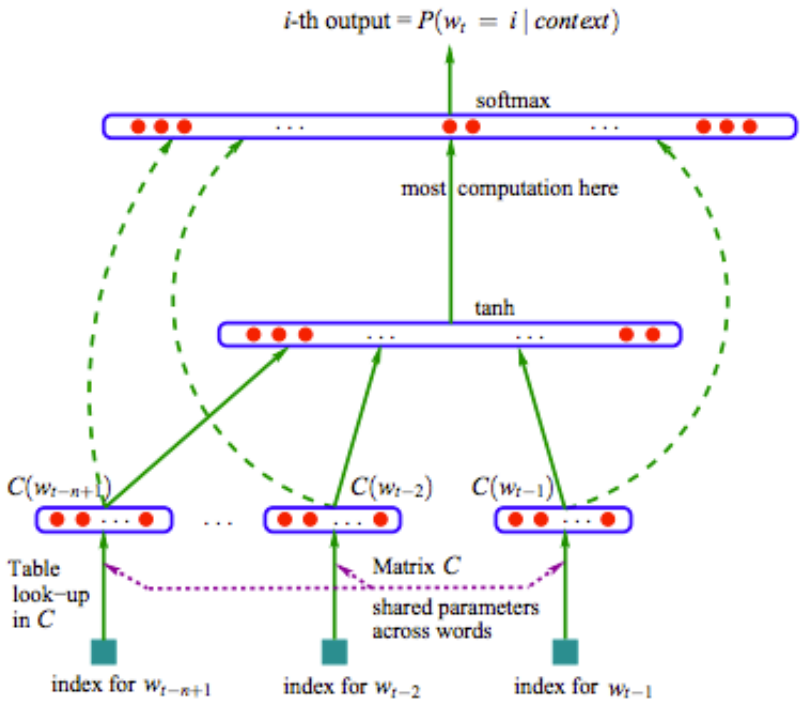
主要参考dl4nlp，做一做assignment

background

word distributed embedding最早是Bengio 03年的论文"A Neural Probabilistic Language Model"提出来,rnn lm 在10年被mikolov提出

use neural networks to model high-dimensional discrete distributions; learn word embedding and probability function at the same time

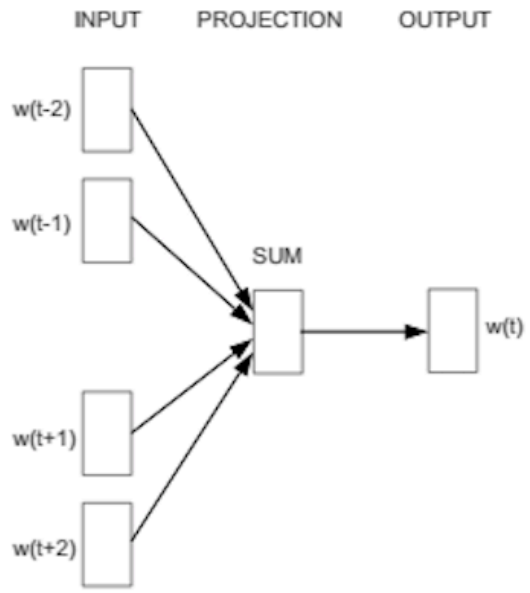
$$P(w_i|w_{t-1}\dots w_1)$$



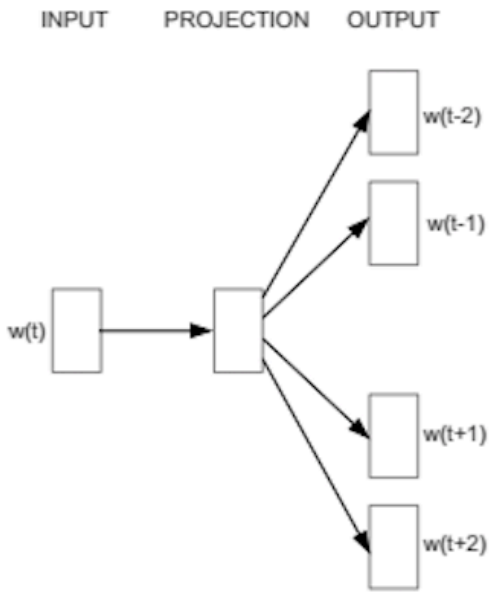
word2vec

主要参考DL4J word2vec

word2vec的思想类似于antodecoder，但是并不是将自身作为训练目标，也不是用RBM来训练。 word2vec将context和word分别作为训练目标，即skip-gram和CBOW。



CBOW



Skip-gram

本质

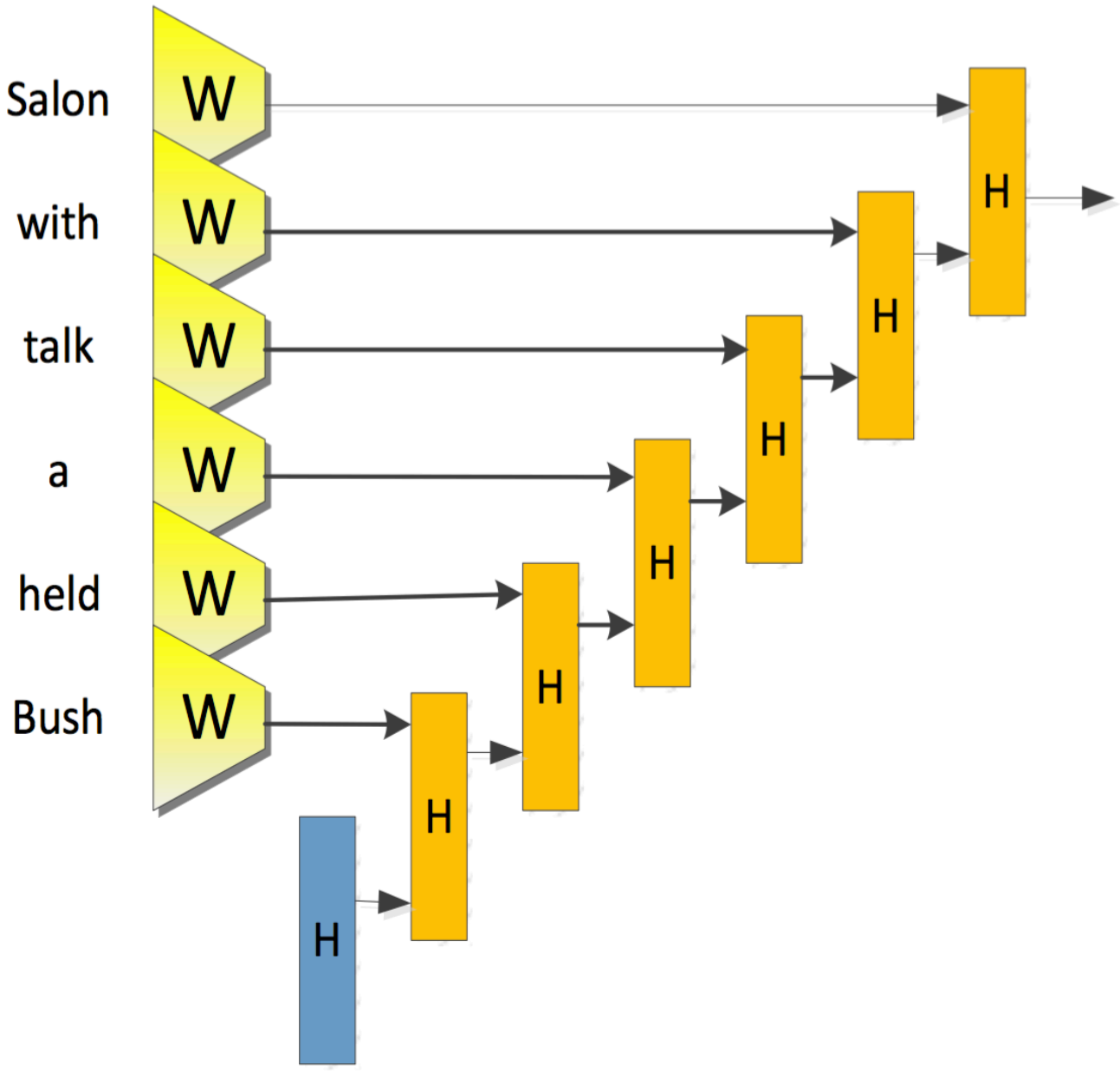
word2vec其实就是two layer shallow neural network,减少了深度神经网络的复杂性，快速的生成word embedding.

- Skip-gram: works well with small amount of the training data, represents well even rare words or phrases.
- CBOW: several times faster to train than the skip-gram, slightly better accuracy for the frequent words

This can get even a bit more complicated if you consider that there are two different ways how to train the models: the normalized hierarchical softmax, and the un-normalized negative sampling. Both work quite differently.

rnn lstm

rnn和lstm应该是在nlp中运用最为广泛的，因为nlp天然需要上下文，而且有很多sequence label的任务，所以这种模型很适合



lstm的原理在colah和Arun的blog中都有非常好的阐述

参考的主要paper是graves的“speech recognition with deep recurrent neural networks”和sequence label那篇，goolge brain的“recurrent neural network regularizaiton”（是tensorflow的lstm的实现方式，高效正则化）

cnn

主要参考Briz用CNN做sentence classification的方法,cnn主要是增强了目标的表达能力，很多人将cnn加入lstm提高表达能力